Minireview

Towards a mutant analysis of the tertiary structures of functional DNA-binding motifs

Andrew Barker, Benno Müller-Hill*

Institut für Genetik der Universität zu Köln, Weyertal 121, D-50931 Köln, Germany

Received 14 May 1998; revised version received 23 June 1998

Abstract Transcription factors have specific regions, often α helices, with which they recognise DNA. These regions are more or less disordered off DNA. Some examples are listed here. However, a detailed mutant analysis of this phenomenon is missing. It could show to what extent DNA binding in vitro and in vivo is harmed when such a region is artificially made rigid by suitable substitutions and could reveal how much transcription factors have improved by having been selected to carry unstable recognition domains.

Key words: Transcription factor; Recognition helix; Induced fit; Unstructured protein region; Cold sensitive mutant; Heat sensitive mutant

1. Introduction

Control of gene expression requires the recognition of specific sites in DNA by specific transcription factors [1]. Even in a simple organism like the bacterium *Escherichia coli*, transcription factors have to be able to recognise a specific sequence in the presence of millions of competing non-specific sites. How is this achieved?

Early work led to the anticipation that protein-DNA interaction would involve preformed, rigid protein surfaces. It therefore came as a surprise when it was demonstrated that some proteins undergo extensive conformational changes upon binding to DNA. As more structural and thermodynamic information has become available, it has become clear that these examples were not exceptions. All DNA-binding proteins that have been suitably analysed to date use this strategy, induced fit, to recognise their binding sites (for a recent review see [2]). These binding sites are buried in a mass of non-specific DNA. In *E. coli* for example a repressor has to scan about four million non-specific DNA binding sites in order to find the one real site. This scanning seems to be more efficient with a disordered DNA binding region. The ordered region may bind too well to non-specific or less specific DNA binding sites, so that the repressor gets lost in the jungle of nonspecific sites and never finds the real, specific site (Fig. 1). However, no relevant genetic analysis of mutants, in which this structural transition has been altered, has been done. We will consider potential outcomes of such an analysis here.

*Corresponding author. Fax: (49) (221) 470-5170.
E-mail: muellerhill@uni-koeln.de

2. Some examples of structural transition

In 1990, it was demonstrated that the basic regions of the bzip proteins Fos and Jun [3], C/EBP [4] and GCN4 [5] are unfolded off their DNA sites. They were shown to bind as α-helices to their specific sites. This was subsequently demonstrated by crystallographic analysis for GCN4 [6,7] and for Jun and Fos [8]. Similarly, the basic regions of bHLH proteins like Max and USF undergo random coil to α-helix folding transitions on binding to the specific DNA recognition sequence 5′-CACGTG [9,10]. The same holds true for the basic region of the SKN-1 transcription factor, where a zipper precedes the basic region [11]. A detailed thermodynamic analysis suggests that all these conformational changes are driven by binding free energy, and therefore are examples of induced fit [12,13]. Thus, there is no overall change in entropy upon binding, and therefore the process does not require an input of energy [12].

This structural change is linked to an interesting property in USF, one of a subset of bHLH proteins in which the bHLH domain is followed by an additional leucine zipper. Addition of this extra zipper domain to a peptide containing the minimal bHLH domain of USF causes a significant increase in the α-helical propensity shown in the absence of DNA [10]. Human USF shows a remarkable cold sensitivity in vitro [14]. Sea urchin USF, which does not contain this extra zipper domain, shows a much reduced cold sensitivity [15]. Domain swapping experiments reveal that it is the presence of the extra leucine zipper in the human USF that is responsible for this increase in cold sensitivity [15].

A comparison of the nuclear magnetic resonance (NMR) structure of the human papillomavirus E2 protein with the X-ray structure of the E2 protein-DNA complex indicates significant differences. In solution, the recognition helix of E2 protein appears to be flexible, as evidenced by fast amide exchange rates. Furthermore, a β-hairpin which contacts DNA is disordered in the NMR structures [16]. A detailed analysis of the stability of the E2 protein suggested that at the very low concentration where DNA binding is normally measured in vitro (e.g. at 10^{-11} M), the E2 protein is predominantly monomeric and unfolded. It thus folds when binding to its DNA target [16,17].

The X-ray and NMR analyses of various homeodomains indicate that they form long, continuous recognition helices when bound to their DNA targets. In the absence of their DNA targets they form the N-terminal recognition helix (helix III); the C-terminal end (helix IV), however, is disordered [18–20].

The amide proton exchange rates of *E. coli* Trp repressor

PII: S0014-5793(98)00797-2

have been measured through their effects on the longitudinal relaxation rates of the amide protons. According to this analysis, the DNA binding region of Trp repressor may be viewed neither as unstructured nor as stable α-helix, but rather as unstable with well-defined helices opening frequently to allow rapid amide proton exchange [21]. This conclusion is supported by various biophysical measurements [22].

The NMR structure of the DNA-binding domain of *E. coli* Lac repressor (the head piece) has been determined [23].The recognition helix of the free head piece does not seem to be significantly disordered compared to the recognition helix of the complex with *lac* operator. However, the conformation of the loop between helix II and helix III changes considerably upon binding to *lac* operator. This change in conformation is essential for binding of the side chains of residues N25 and H29 to *lac* operator DNA. In addition, the head piece domain is not sufficiently rigid to be visible in X-ray analyses of Lac repressor crystals in the absence of DNA [24].

Is the phenomenon of induced fit unique to DNA-binding proteins that recognise DNA with an α-helix? An analysis of the bacteriophage P22 Arc repressor indicates otherwise. Arc repressor recognises its target sequence with a β-sheet [25]. A mutation which increases the stability of the protein by reducing the rate of protein unfolding, also reduces operator binding drastically in vitro, and abolishes repression in vivo [26]. In other words, the need to be disordered off DNA is not restricted to α-helical proteins.

Recent reports also indicate that the α-helical activation domains of two eukaryotic transcription factors, c-Myc [27] and VP16 [28], use a similar strategy to interact with TATA-box binding proteins: both are disordered in the absence of their targets. Thus it seems that transcription factors have used the same answer to two very different problems involving recognition of macromolecular partners. Furthermore, a recent report has shown that induced fit is important for recognition of the RNA target of the bacteriophage λ antitermination protein N. The N-terminal domain of the protein forms an α-helix when bound to its target, but off RNA the entire protein is disordered [29].

3. Genetic analysis

To understand why these structures are only formed on DNA, let us consider possible mutants in this flexible region. We will specifically consider a protein which uses an α-helix to bind DNA in *E. coli*, but bear in mind that the outcome should be similar for other classes of protein and other organisms. There are two modes for disturbing the DNA-binding region such that it may become non-functional: (i) its α-helical tendency may be weakened to such an extent that it may have difficulty in forming an α-helix even in the presence of the DNA site. A subset of such mutants should be heat sensitive. At 30°C (or lower) they may still bind to their DNA target, but at 40°C they may be unable to do so, both in vivo and in vitro. (ii) The recognition helix may become a rigid helix off DNA. A subset of such mutants should be cold sensitive. This may then be disadvantageous for operator binding at 30°C but not at 40°C.

This is of course an oversimplification. An effective cold sensitive mutation may not lead directly to an increase of helix propensity. There may be indirect cooperative effects on residues directly interacting with the DNA, with other parts of the same protein or with other proteins. So far, mutants of transcription factors have not been selected, or designed, in which the recognition helix is deliberately stabilised. But do such mutants already exist? We found two possible examples in the literature. The first is an analysis of the effects that mutations, which lead to specificity changes in GCN4 peptides in vitro, have when introduced into full-length GCN4 in vivo. Astonishingly, most of the observed effects on binding are very different in vitro and in vivo [30]. The authors suggest that this result might be due to different bending of target DNA by full-length GCN4 in vivo, and GCN4 peptides in vitro. It seems likely that changing the propensity of the basic region to form an α-helix may also be involved, at least in some cases. For example, mutant A-14S GCN4, the peptide form of which binds a particular target DNA at 4°C in vitro, does not seem to bind as full-length protein to the same target at 30°C in vivo [30]. This may be a mutant of the first, heat

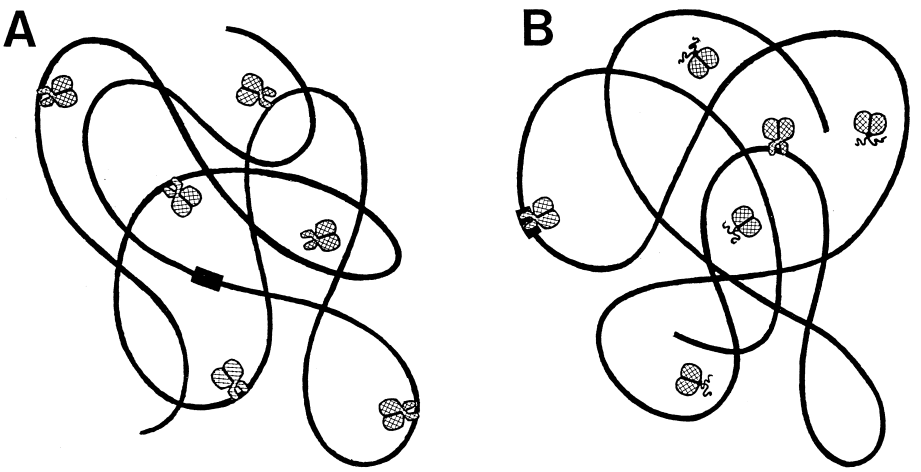

Fig. 1. Predicted properties of a cold sensitive repressor mutant in which the DNA binding domain is rigid off DNA at low temperature. A: Low temperature. The DNA binding motif is rigid off DNA. The dimeric repressor binds with high frequency to non-specific DNA sites. This leads to non-occupation of the specific DNA binding site (black box). B: High temperature. The DNA binding motif is disordered off DNA. The DNA binding motif becomes rigid and ordered when bound to DNA. The dimeric repressor binds preferably to its specific binding site (black box) in vivo.

sensitive, type. On the other hand, the peptide form of mutant A-14C does not bind to a particular target at 4°C in vitro, whereas the full-length form binds to the same target at 30°C in vivo [30]. This may be interpreted as a cold sensitive mutant, which is rigid at low temperature but flexible at higher temperature. We are well aware that other explanations are possible. Time may show.

Cold sensitive mutants have been described in Lac repressor. Of the four thousand Lac repressor mutants analysed by Jeffrey Miller and his group [31], thirteen are cold sensitive. Eleven of these cold sensitive mutants occur in the head piece between residue 16 (in the recognition helix) and residue 35 (in the helix following the recognition helix). These are the mutants: S16C, A; V20C; V23C, A; H29A, E; K33H, F, E; and R35E. It can be seen that in most of the mutants a residue of higher α-helical propensity has been introduced. Two positions (33 and 35) are solvent exposed and without function. The others are involved in DNA binding or stabilisation of tertiary structure and are therefore difficult to interpret. The replacements which lead to the cold sensitive phenotype in residues 33 and 35 are all of higher helical propensity. A detailed analysis of these mutants offers a test of our hypothesis.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 274.

References

[1] McKnight, S.L. and Yamamoto, K.R. (Eds.) (1992) Transcriptional Regulation (2 Volumes), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[2] Lefstin, J.A. and Yamamoto, K.R. (1998) Nature 392, 885–888.
[3] Patel, L., Abate, C. and Curran, T. (1990) Nature 347, 572–575.
[4] Shuman, J.D., Vinson, C.R. and McKnight, S.L. (1990) Science 249, 771–774.
[5] O'Neil, K.T., Hoess, R.H. and DeGrado, W.F. (1990) Science 249, 774–778.
[6] Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) Cell 71, 1223–1237.
[7] König, P. and Richmond, T.J. (1993) J. Mol. Biol. 233, 139–154.
[8] Chen, L., Glover, J.N.M., Hogan, P.G., Rao, A. and Harrison, S.C. (1998) Nature 392, 42–48.
[9] Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B. and Burley, S.K. (1993) Nature 363, 38–45.
[10] Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G. and Burley, S.K. (1994) EMBO J. 13, 180–189.
[11] Carroll, A.S., Gilbert, D.E., Liu, X., Cheung, J.W., Michnowicz, J.E., Wagner, G., Ellenberger, T.E. and Blackwell, T.K. (1997) Genes Dev. 11, 2227–2238.
[12] Spolar, R.S. and Record Jr., M.T. (1994) Science 263, 777–784.
[13] Burley, S.K. (1994) Nat. Struct. Biol. 1, 207–208.
[14] Bresnick, E.H. and Felsenfeld, G. (1993) J. Biol. Chem. 268, 18824–18834.
[15] Lu, T. and Sawadogo, M. (1994) J. Biol. Chem. 269, 30694–30700.
[16] Liang, H., Petros, A.M., Meadows, R.P., Yoon, H.S., Egan, D.A., Walter, K., Holzman, T.F., Robins, T. and Fesik, S.W. (1996) Biochemistry 35, 2095–2103.
[17] Mok, Y.K., dePrat, G.G., Butler, P.J. and Bycroft, M. (1996) Protein Sci. 5, 310–319.
[18] Gehring, W.J., Quian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resedenz-Perez, D., Affolter, M., Otting, G. and Wüthrich, K. (1994) Cell 78, 211–223.
[19] Sivaraja, M., Botfield, M.C., Mueller, M., Jansco, A. and Weiss, M.A. (1994) Biochemistry 33, 9845–9855.
[20] Tsao, D.H., Gruschus, J.M., Wang, L.H., Nirenberg, M. and Ferretti, J.A. (1995) J. Mol. Biol. 251, 297–307.
[21] Gryk, M.R., Finucane, M.D., Zheng, Z. and Jardetzky, O. (1995) J. Mol. Biol. 246, 618–627.
[22] Reedstrom, R.M. and Royer, C.A. (1995) J. Mol. Biol. 253, 266–276.
[23] Slijper, M., Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1996) J. Mol. Biol. 259, 761–773.
[24] Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. and Lu, P. (1996) Science 271, 1247–1254.
[25] Raumann, B.E., Rould, M.A., Pabo, C.O. and Sauer, R.T. (1994) Nature 367, 754–757.
[26] Schildbach, J.F., Milla, M.E., Jeffrey, P.D., Raumann, B.E. and Sauer, R.T. (1995) Biochemistry 34, 1405–1412.
[27] McEwan, I.J., Dahlman-Wright, K., Ford, J. and Wright, A.P.H. (1996) Biochemistry 35, 9584–9593.
[28] Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J. and Verdine, G.L. (1997) Science 277, 1310–1313.
[29] Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. (1998) Cell 93, 289–299.
[30] Suckow, M. and Hollenberg, C.P. (1998) J. Mol. Biol. 259, 761–773.
[31] Suckow, J., Markiewicz, P., Kleina, L.G., Miller, J., Kisters-Woike, B. and Müller-Hill, B. (1996) J. Mol. Biol. 261, 509–523.